

NOVEL METHODS OF CONSTRUCTING LIBRARIES OF GENETIC
PACKAGES THAT COLLECTIVELY DISPLAY THE MEMBERS OF A
DIVERSE FAMILY OF PEPTIDES, POLYPEPTIDES OR PROTEINS

5 The present invention relates to constructing
libraries of genetic packages that display a member of
a diverse family of peptides, polypeptides or proteins
and collectively display at least a portion of the
diversity of the family. In a preferred embodiment,
the displayed polypeptides are human Fabs.

10 More specifically, the invention is directed
to the methods of cleaving single-stranded nucleic
acids at chosen locations, the cleaved nucleic acids
encoding, at least in part, the peptides, polypeptides
or proteins displayed on the genetic packages of the
15 libraries of the invention. In a preferred embodiment,
the genetic packages are filamentous phage or
phagemids.

The present invention further relates to
methods of screening the libraries of genetic packages
20 that display useful peptides, polypeptides and proteins
and to the peptides, polypeptides and proteins
identified by such screening.

BACKGROUND OF THE INVENTION

It is now common practice in the art to prepare libraries of genetic packages that display a member of a diverse family of peptides, polypeptides or proteins and collectively display at least a portion of the diversity of the family. In many common libraries, the displayed peptides, polypeptides or proteins are related to antibodies. Often, they are Fabs or single chain antibodies.

10 In general, the DNAs that encode members of the families to be displayed must be amplified before they are cloned and used to display the desired member on the surface of a genetic package. Such amplification typically makes use of forward and
15 backward primers.

Such primers can be complementary to sequences native to the DNA to be amplified or complementary to oligonucleotides attached at the 5' or 3' ends of that DNA. Primers that are complementary to
20 sequences native to the DNA to be amplified are disadvantaged in that they bias the members of the families to be displayed. Only those members that contain a sequence in the native DNA that is substantially complementary to the primer will be
25 amplified. Those that do not will be absent from the family. For those members that are amplified, any diversity within the primer region will be suppressed.

For example, in European patent 368,684 B1, the primer that is used is at the 5' end of the V_H
30 region of an antibody gene. It anneals to a sequence region in the native DNA that is said to be "sufficiently well conserved" within a single species. Such primer will bias the members amplified to those

having this "conserved" region. Any diversity within this region is extinguished.

It is generally accepted that human antibody genes arise through a process that involves a combinatorial selection of V and J or V, D, and J followed by somatic mutations. Although most diversity occurs in the Complementary Determining Regions (CDRs), diversity also occurs in the more conserved Framework Regions (FRs) and at least some of this diversity confers or enhances specific binding to antigens (Ag). As a consequence, libraries should contain as much of the CDR and FR diversity as possible.

To clone the amplified DNAs for display on a genetic package of the peptides, polypeptides or proteins that they encode, the DNAs must be cleaved to produce appropriate ends for ligation to a vector. Such cleavage is generally effected using restriction endonuclease recognition sites carried on the primers. When the primers are at the 5' end of DNA produced from reverse transcription of RNA, such restriction leaves deleterious 5' untranslated regions in the amplified DNA. These regions interfere with expression of the cloned genes and thus the display of the peptides, polypeptides and proteins coded for by them.

SUMMARY OF THE INVENTION

It is an object of this invention to provide novel methods for constructing libraries of genetic packages that display a member of a diverse family of peptides, polypeptides or proteins and collectively display at least a portion of the diversity of the family. These methods are not biased toward DNAs that contain native sequences that are complementary to the

primers used for amplification. They also enable any sequences that may be deleterious to expression to be removed from the amplified DNA before cloning and displaying.

5 It is another object of this invention to provide a method for cleaving single-stranded nucleic acid sequences at a desired location, the method comprising the steps of:

10 (i) contacting the nucleic acid with a single-stranded oligonucleotide, the oligonucleotide being functionally complementary to the nucleic acid in the region in which cleavage is desired and including a sequence that with its complement
15 in the nucleic acid forms a restriction endonuclease recognition site that on restriction results in cleavage of the nucleic acid at the desired location; and

20 (ii) cleaving the nucleic acid solely at the recognition site formed by the complementation of the nucleic acid and the oligonucleotide;

the contacting and the cleaving steps being performed at a temperature sufficient to maintain the nucleic
25 acid in substantially single-stranded form, the oligonucleotide being functionally complementary to the nucleic acid over a large enough region to allow the two strands to associate such that cleavage may occur at the chosen temperature and at the desired location,
30 and the cleavage being carried out using a restriction endonuclease that is active at the chosen temperature.

It is a further object of this invention to provide an alternative method for cleaving single-

stranded nucleic acid sequences at a desired location,
the method comprising the steps of:

- 5 (i) contacting the nucleic acid with a
partially double-stranded oligonucleotide,
the single-stranded region of the
oligonucleotide being functionally
complementary to the nucleic acid in the
region in which cleavage is desired, and the
double-stranded region of the oligonucleotide
10 having a Type II-S restriction endonuclease
recognition site, whose cleavage site is
located at a known distance from the
recognition site; and
- 15 (ii) cleaving the nucleic acid solely at
the cleavage site formed by the
complementation of the nucleic acid and the
single-stranded region of the
oligonucleotide;

the contacting and the cleaving steps being performed
20 at a temperature sufficient to maintain the nucleic
acid in substantially single-stranded form, the
oligonucleotide being functionally complementary to the
nucleic acid over a large enough region to allow the
two strands to associate such that cleavage may occur
25 at the chosen temperature and at the desired location,
and the cleavage being carried out using a restriction
endonuclease that is active at the chosen temperature.

It is another objective of the present
invention to provide a method of capturing DNA
30 molecules that comprise a member of a diverse family of
DNAs and collectively comprise at least a portion of
the diversity of the family. These DNA molecules in
single-stranded form have been cleaved by one of the

methods of this invention. This method involves ligating the individual single-stranded DNA members of the family to a partially duplex DNA complex. The method comprises the steps of:

5 (i) contacting a single-stranded nucleic acid sequence that has been cleaved with a restriction endonuclease with a partially double-stranded oligonucleotide, the single-stranded region of the oligonucleotide being
10 functionally complementary to the nucleic acid in the region that remains after cleavage, the double-stranded region of the oligonucleotide including any sequences necessary to return the sequences that remain
15 after cleavage into proper reading frame for expression and containing a restriction endonuclease recognition site 5' of those sequences; and

20 (ii) cleaving the partially double-stranded oligonucleotide sequence solely at the restriction endonuclease recognition site contained within the double-stranded region of the partially double-stranded oligonucleotide.

25 It is another object of this invention to prepare libraries, that display a diverse family of peptides, polypeptides or proteins and collectively display at least part of the diversity of the family, using the methods and DNAs described above.

30 It is an object of this invention to screen those libraries to identify useful peptides, polypeptides and proteins and to use those substances in human therapy.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic of various methods that may be employed to amplify VH genes without using primers specific for VH sequences.

5 FIG. 2 is a schematic of various methods that may be employed to amplify VL genes without using VL sequences.

FIG. 3 depicts gel analysis of cleaved kappa DNA from Example 2.

10 FIG. 4 depicts gel analysis of cleaved kappa DNA from Example 2.

FIG. 5 depicts gel analysis of amplified kappa DNA from Example 2.

15 FIG. 6 depicts gel purified amplified kappa DNA from Example 2.

TERMS

In this application, the following terms and abbreviations are used:

20	Sense strand	The upper strand of ds DNA as usually written. In the sense strand, 5'-ATG-3' codes for Met.
25	Antisense strand	The lower strand of ds DNA as usually written. In the antisense strand, 3'-TAC-5' would correspond to a Met codon in the sense strand.

	Forward primer:	A "forward" primer is complementary to a part of the sense strand and primes for synthesis of a new antisense-strand molecule. "Forward primer" and "lower-strand primer" are equivalent.
5		
	Backward primer:	A "backward" primer is complementary to a part of the antisense strand and primes for synthesis of a new sense-strand molecule. "Backward primer" and "top-strand primer" are equivalent.
10		
15	Bases:	Bases are specified either by their position in a vector or gene as their position within a gene by codon and base. For example, "89.1" is the first base of codon 89, 89.2 is the second base of codon 89.
20		
	Sv	Streptavidin
	Ap	Ampicillin
	ap ^R	A gene conferring ampicillin resistance.
25		
	RE	Restriction endonuclease

	URE	Universal restriction endonuclease
5	Functionally complementary	Two sequences are sufficiently complementary so as to anneal under the chosen conditions.
	RERS	Restriction endonuclease recognition site
	AA	Amino acid
10	PCR	Polymerization chain reaction
	GLGs	Germline genes
15	Ab	Antibody: an immunoglobulin. The term also covers any protein having a binding domain which is homologous to an immunoglobulin binding domain. A few examples of antibodies within this definition are, <i>inter alia</i> , immunoglobulin isotypes and the Fab, F(ab ¹) ₂ , scfv, Fv, dAb and Fd fragments.
20		
25	Fab	Two chain molecule comprising an Ab light chain and part of a heavy-chain.

	scFv	A single-chain Ab comprising either VH::linker::VL or VL::linker::VH
	w.t.	Wild type
5	HC	Heavy chain
	LC	Light chain
	VK	A variable domain of a Kappa light chain.
10	VH	A variable domain of a heavy chain.
	VL	A variable domain of a lambda light chain.

In this application, all references referred to are specifically incorporated by reference.

15 DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

 The nucleic acid sequences that are useful in the methods of this invention, i.e., those that encode at least in part the individual peptides, polypeptides and proteins displayed on the genetic packages of this invention, may be naturally occurring, synthetic or a combination thereof. They may be mRNA, DNA or cDNA. In the preferred embodiment, the nucleic acids encode antibodies. Most preferably, they encode Fabs.

 The nucleic acids useful in this invention may be naturally diverse, synthetic diversity may be

introduced into those naturally diverse members, or the diversity may be entirely synthetic. For example, synthetic diversity can be introduced into one or more CDRs of antibody genes.

5 Synthetic diversity may be created, for example, through the use of TRIM technology (U.S. 5,869,644). TRIM technology allows control over exactly which amino-acid types are allowed at variegated positions and in what proportions. In TRIM
10 technology, codons to be diversified are synthesized using mixtures of trinucleotides. This allows any set of amino acid types to be included in any proportion.

 Another alternative that may be used to generate diversified DNA is mixed oligonucleotide
15 synthesis. With TRIM technology, one could allow Ala and Trp. With mixed oligonucleotide synthesis, a mixture that included Ala and Trp would also necessarily include Ser and Gly. The amino-acid types allowed at the variegated positions are picked with
20 reference to the structure of antibodies, or other peptides, polypeptides or proteins of the family, the observed diversity in germline genes, the observed somatic mutations frequently observed, and the desired areas and types of variegation.

25 In a preferred embodiment of this invention, the nucleic acid sequences for at least one CDR or other region of the peptides, polypeptides or proteins of the family are cDNAs produced by reverse transcription from mRNA. More preferably, the mRNAs
30 are obtained from peripheral blood cells, bone marrow cells, spleen cells or lymph node cells (such as B-lymphocytes or plasma cells) that express members of naturally diverse sets of related genes. More preferable, the mRNAs encode a diverse family of

antibodies. Most preferably, the mRNAs are obtained from patients suffering from at least one autoimmune disorder or cancer. Preferably, mRNAs containing a high diversity of autoimmune diseases, such as systemic
5 lupus erythematosus, systemic sclerosis, rheumatoid arthritis, antiphospholipid syndrome and vasculitis are used.

In a preferred embodiment of this invention, the cDNAs are produced from the mRNAs using reverse
10 transcription. In this preferred embodiment, the mRNAs are separated from the cell and degraded using standard methods, such that only the full length (i.e., capped) mRNAs remain. The cap is then removed and reverse transcription used to produce the cDNAs.

15 The reverse transcription of the first (antisense) strand can be done in any manner with any suitable primer. See, e.g., HJ de Haard et al., Journal of Biological Chemistry, 274(26):18218-30 (1999). In the preferred embodiment of this invention
20 where the mRNAs encode antibodies, primers that are complementary to the constant regions of antibody genes may be used. Those primers are useful because they do not generate bias toward subclasses of antibodies. In another embodiment, poly-dT primers may be used (and
25 may be preferred for the heavy-chain genes). Alternatively, sequences complementary to the primer may be attached to the termini of the antisense strand.

In one preferred embodiment of this invention, the reverse transcriptase primer may be
30 biotinylated, thus allowing the cDNA product to be immobilized on streptavidin (Sv) beads. Immobilization can also be effected using a primer labeled at the 5' end with one of a) free amine group, b) thiol, c) carboxylic acid, or d) another group not found in DNA

that can react to form a strong bond to a known partner on an insoluble medium. If, for example, a free amine (preferably primary amine) is provided at the 5' end of a DNA primer, this amine can be reacted with carboxylic acid groups on a polymer bead using standard amide-forming chemistry. If such preferred immobilization is used during reverse transcription, the top strand RNA is degraded using well-known enzymes, such as a combination of RNaseH and RNaseA, either before or after immobilization.

The nucleic acid sequences useful in the methods of this invention are generally amplified before being used to display the peptides, polypeptides or proteins that they encode. Prior to amplification, the single-stranded DNAs may be cleaved using either of the methods described before. Alternatively, the single-stranded DNAs may be amplified and then cleaved using one of those methods.

Any of the well known methods for amplifying nucleic acid sequences may be used for such amplification. Methods that maximize, and do not bias, diversity are preferred. In a preferred embodiment of this invention where the nucleic acid sequences are derived from antibody genes, the present invention preferably utilizes primers in the constant regions of the heavy and light chain genes and primers to a synthetic sequence that are attached at the 5' end of the sense strand. Priming at such synthetic sequence avoids the use of sequences within the variable regions of the antibody genes. Those variable region priming sites generate bias against V genes that are either of rare subclasses or that have been mutated at the priming sites. This bias is partly due to suppression of diversity within the primer region and partly due to

lack of priming when many mutations are present in the region complementary to the primer. The methods disclosed in this invention have the advantage of not biasing the population of amplified antibody genes for particular V gene types.

The synthetic sequences may be attached to the 5' end of the DNA strand by various methods well known for ligating DNA sequences together. RT CapExtension is one preferred method.

10 In RT CapExtension (derived from Smart PCR^(TM)), a short overlap (5'-...GGG-3' in the upper-strand primer (USP-GGG) complements 3'-CCC....5' in the lower strand) and reverse transcriptases are used so that the reverse complement of the upper-strand primer
15 is attached to the lower strand.

In a preferred embodiment of this invention, the upper strand or lower strand primer may be also biotinylated or labeled at the 5' end with one of a) free amino group, b) thiol, c) carboxylic acid and d)
20 another group not found in DNA that can react to form a strong bond to a known partner as an insoluble medium. These can then be used to immobilize the labeled strand after amplification. The immobilized DNA can be either single or double-stranded.

25 FIG. 1 shows a schematic of the amplification of VH genes. FIG. 1, Panel A shows a primer specific to the poly-dT region of the 3' UTR priming synthesis of the first, lower strand. Primers that bind in the constant region are also suitable. Panel B shows the
30 lower strand extended at its 3' end by three Cs that are not complementary to the mRNA. Panel C shows the result of annealing a synthetic top-strand primer ending in three GGGs that hybridize to the 3' terminal CCCs and extending the reverse transcription extending

the lower strand by the reverse complement of the synthetic primer sequence. Panel D shows the result of PCR amplification using a 5' biotinylated synthetic top-strand primer that replicates the 5' end of the synthetic primer of panel C and a bottom-strand primer complementary to part of the constant domain. Panel E shows immobilized double-stranded (ds) cDNA obtained by using a 5'-biotinylated top-strand primer.

FIG. 2 shows a similar schematic for amplification of VL genes. FIG. 2, Panel A shows a primer specific to the constant region at or near the 3' end priming synthesis of the first, lower strand. Primers that bind in the poly-dT region are also suitable. Panel B shows the lower strand extended at its 3' end by three Cs that are not complementary to the mRNA. Panel C shows the result of annealing a synthetic top-strand primer ending in three GGGs that hybridize to the 3' terminal CCCs and extending the reverse transcription extending the lower strand by the reverse complement of the synthetic primer sequence. Panel D shows the result of PCR amplification using a 5' biotinylated synthetic top-strand primer that replicates the 5' end of the synthetic primer of panel C and a bottom-strand primer complementary to part of the constant domain. The bottom-strand primer also contains a useful restriction endonuclease site, such as AscI. Panel E shows immobilized ds cDNA obtained by using a 5'-biotinylated top-strand primer.

In FIGs. 1 and 2, each V gene consists of a 5' untranslated region (UTR) and a secretion signal, followed by the variable region, followed by a constant region, followed by a 3' untranslated region (which typically ends in poly-A). An initial primer for reverse transcription may be complementary to the

constant region or to the poly A segment of the 3'-UTR. For human heavy-chain genes, a primer of 15 T is preferred. Reverse transcriptases attach several C residues to the 3' end of the newly synthesized DNA.

5 RT CapExtension exploits this feature. The reverse transcription reaction is first run with only a lower-strand primer. After about 1 hour, a primer ending in GGG (USP-GGG) and more RTase are added. This causes the lower-strand cDNA to be extended by the reverse
10 complement of the USP-GGG up to the final GGG. Using one primer identical to part of the attached synthetic sequence and a second primer complementary to a region of known sequence at the 3' end of the sense strand, all the V genes are amplified irrespective of their V
15 gene subclass.

After amplification, the DNAs of this invention are rendered single-stranded. For example, the strands can be separated by using a biotinylated primer, capturing the biotinylated product on
20 streptavidin beads, denaturing the DNA, and washing away the complementary strand. Depending on which end of the captured DNA is wanted, one will choose to immobilize either the upper (sense) strand or the lower (antisense) strand.

25 To prepare the single-stranded amplified DNAs for cloning into genetic packages so as to effect display of the peptides, polypeptides or proteins encoded, at least in part, by those DNAs, they must be manipulated to provide ends suitable for cloning and
30 expression. In particular, any 5' untranslated regions and mammalian signal sequences must be removed and replaced, in frame, by a suitable signal sequence that functions in the display host. Additionally, parts of the variable domains (in antibody genes) may be removed

and replaced by synthetic segments containing synthetic diversity. The diversity of other gene families may likewise be expanded with synthetic diversity.

According to the methods of this invention,
5 there are two ways to manipulate the single-stranded amplified DNAs for cloning. The first method comprises the steps of:

10 (i) contacting the nucleic acid with a single-stranded oligonucleotide, the oligonucleotide being functionally complementary to the nucleic acid in the region in which cleavage is desired and including a sequence that with its complement in the nucleic acid forms a restriction
15 endonuclease recognition site that on restriction results in cleavage of the nucleic acid at the desired location; and

20 (ii) cleaving the nucleic acid solely at the recognition site formed by the complementation of the nucleic acid and the oligonucleotide;

the contacting and the cleaving steps being performed at a temperature sufficient to maintain the nucleic acid in substantially single-stranded form, the
25 oligonucleotide being functionally complementary to the nucleic acid over a large enough region to allow the two strands to associate such that cleavage may occur at the chosen temperature and at the desired location, and the cleavage being carried out using a restriction
30 endonuclease that is active at the chosen temperature.

In this first method, short oligonucleotides are annealed to the single-stranded DNA so that restriction endonuclease recognition sites formed

within the now locally double-stranded regions of the DNA can be cleaved. In particular, a recognition site that occurs at the same position in a substantial fraction of the single-stranded DNAs is identical.

5 For antibody genes, this can be done using a catalog of germline sequences. See, e.g.,
"http://www.mrc-cpe.cam.ac.uk/imt-doc/restricted/ok.html." Updates can be obtained from this site under the heading "Amino acid and nucleotide sequence
10 alignments." For other families, similar comparisons exist and may be used to select appropriate regions for cleavage and to maintain diversity.

For example, Table 195 depicts the DNA sequences of the FR3 regions of the 51 known human VH
15 germline genes. In this region, the genes contain restriction endonuclease recognition sites shown in Table 200. Restriction endonucleases that cleave a large fraction of germline genes at the same site are preferred over endonucleases that cut at a variety of
20 sites. Furthermore, it is preferred that there be only one site for the restriction endonucleases within the region to which the short oligonucleotide binds on the single-stranded DNA, e.g., about 10 bases on either side of the restriction endonuclease recognition site.

25 An enzyme that cleaves downstream in FR3 is also more preferable because it captures fewer mutations in the framework. This may be advantageous in some cases. However, it is well known that framework mutations exist and confer and enhance
30 antibody binding. The present invention, by choice of appropriate restriction site, allows all or part of FR3 diversity to be captured. Hence, the method also allows extensive diversity to be captured.

Finally, in the methods of this invention restriction endonucleases that are active between about 45° and about 75°C are used. Preferably enzymes that are active above 50°C, and more preferably active about 5 55°C, are used. Such temperatures maintain the nucleic acid sequence to be cleaved in substantially single-stranded form.

Enzymes shown in Table 200 that cut many of the heavy chain FR3 germline genes at a single position 10 include: *MaeIII*(24@4), *Tsp45I*(21@4), *HphI*(44@5), *BsaJI*(23@65), *AluI*(23@47), *BlpI*(21@48), *DdeI*(29@58), *BglII*(10@61), *MslI*(44@72), *BsiEI*(23@74), *EaeI*(23@74), *EagI*(23@74), *HaeIII*(25@75), *Bst4CI*(51@86), *HpyCH4III*(51@86), *HinfI*(38@2), *MlyI*(18@2), *PleI*(18@2), 15 *MnlI*(31@67), *HpyCH4V*(21@44), *BsmAI*(16@11), *BpmI*(19@12), *XmnI*(12@30), and *SacI*(11@51). (The notation used means, for example, that *BsmAI* cuts 16 of the FR3 germline genes with a restriction endonuclease recognition site beginning at base 11 of FR3.)

20 For cleavage of human heavy chains in FR3, the preferred restriction endonucleases are: *Bst4CI* (or *TaaI* or *HpyCH4III*), *BlpI*, *HpyCH4V*, and *MslI*. Because ACNGT (the restriction endonuclease recognition site for *Bst4CI*, *TaaI*, and *HpyCH4III*) is found at a 25 consistent site in all the human FR3 germline genes, one of those enzymes is the most preferred for capture of heavy chain CDR3 diversity. *BlpI* and *HpyCH4V* are complementary. *BlpI* cuts most members of the VH1 and VH4 families while *HpyCH4V* cuts most members of the 30 VH3, VH5, VH6, and VH7 families. Neither enzyme cuts VH2s, but this is a very small family, containing only three members. Thus, these enzymes may also be used in preferred embodiments of the methods of this invention.

The restriction endonucleases *HpyCH4III*,
Bst4CI, and *TaaI* all recognize 5'-ACnGT-3' and cut
upper strand DNA after n and lower strand DNA before
the base complementary to n. This is the most
5 preferred restriction endonuclease recognition site for
this method on human heavy chains because it is found
in all germline genes. Furthermore, the restriction
endonuclease recognition region (ACnGT) matches the
second and third bases of a tyrosine codon (tay) and
10 the following cysteine codon (tg~~y~~) as shown in Table
206. These codons are highly conserved, especially the
cysteine in mature antibody genes.

Table 250 E shows the distinct
oligonucleotides of length 22 (except the last one
15 which is of length 20) bases. Table 255 C shows the
analysis of 1617 actual heavy chain antibody genes. Of
these, 1511 have the site and match one of the
candidate oligonucleotides to within 4 mismatches.
Eight oligonucleotides account for most of the matches
20 and are given in Table 250 F.1. The 8 oligonucleotides
are very similar so that it is likely that satisfactory
cleavage will be achieved with only one oligonucleotide
(such as H43.77.97.1-02#1) by adjusting temperature,
pH, salinity, and the like. One or two
25 oligonucleotides may likewise suffice whenever the
germline gene sequences differ very little and
especially if they differ very little close to the
restriction endonuclease recognition region to be
cleaved. Table 255 D shows a repeat analysis of 1617
30 actual heavy chain antibody genes using only the 8
chosen oligonucleotides. This shows that 1463 of the
sequences match at least one of the oligonucleotides to
within 4 mismatches and have the site as expected.

Only 7 sequences have a second *Hpy*CH4III restriction endonuclease recognition region in this region.

Another illustration of choosing an appropriate restriction endonuclease recognition site involves cleavage in FR1 of human heavy chains. Cleavage in FR1 allows capture of the entire CDR diversity of the heavy chain.

The germline genes for human heavy chain FR1 are shown in Table 217. Table 220 shows the restriction endonuclease recognition sites found in human germline genes FR1s. The preferred sites are *Bsg*I (GTGCAG;39@4), *Bso*FI (GCngc;43@6,11@9,2@3,1@12), *Tse*I (Gcwgc;43@6,11@9,2@3,1@12), *Msp*AI (CMGckg;46@7,2@1), *Pvu*II (CAGctg;46@7,2@1), *Alu*I (AGct;48@82@2), *Dde*I (Ctnag;22@52,9@48), *Hph*I (tcacc;22@80), *Bss*KI (Nccngg;35@39,2@40), *Bsa*JI (Ccnngg;32@40,2@41), *Bst*NI (CCwgg;33@40), *Scr*FI (CCngg;35@40,2@41), *Eco*O109I (RGgnccy;22@46,11@43), *Sau*96I (Ggncc;23@47,11@44), *Ava*II (Ggwcc;23@47,4@44), *Ppu*MI (RGgwccy;22@46,4@43), *Bsm*FI (gtccc;20@48), *Hin*FI (Gantc;34@16,21@56,21@77), *Tfi*I (21@77), *Mly*I (GAGTC;34@16), *Mly*I (gactc;21@56), and *Alw*NI (CAGnnnctg;22@68). The more preferred sites are *Msp*AI and *Pvu*II. *Msp*AI and *Pvu*II have 46 sites at 7-12 and 2 at 1-6. To avoid cleavage at both sites, oligonucleotides are used that do not fully cover the site at 1-6. Thus, the DNA will not be cleaved at that site. We have shown that DNA that extends 3, 4, or 5 bases beyond a *Pvu*II-site can be cleaved efficiently.

Another illustration of choosing an appropriate restriction endonuclease recognition site involves cleavage in FR1 of human kappa light chains. Table 300 shows the human kappa FR1 germline genes and

Table 302 shows restriction endonuclease recognition sites that are found in a substantial number of human kappa FR1 germline genes at consistent locations. Of the restriction endonuclease recognition sites listed, 5 *Bsm*AI and *Pfl*FI are the most preferred enzymes. *Bsm*AI sites are found at base 18 in 35 of 40 germline genes. *Pfl*FI sites are found in 35 of 40 germline genes at base 12.

Another example of choosing an appropriate 10 restriction endonuclease recognition site involves cleavage in FR1 of the human lambda light chain. Table 400 shows the 31 known human lambda FR1 germline gene sequences. Table 405 shows restriction endonuclease recognition sites found in human lambda FR1 germline 15 genes. *Hinf*I and *Dde*I are the most preferred restriction endonucleases for cutting human lambda chains in FR1.

After the appropriate site or sites for cleavage are chosen, one or more short oligonucleotides 20 are prepared so as to functionally complement, alone or in combination, the chosen recognition site. The oligonucleotides also include sequences that flank the recognition site in the majority of the amplified genes. This flanking region allows the sequence to 25 anneal to the single-stranded DNA sufficiently to allow cleavage by the restriction endonuclease specific for the site chosen.

The actual length and sequence of the oligonucleotide depends on the recognition site and the 30 conditions to be used for contacting and cleavage. The length must be sufficient so that the oligonucleotide is functionally complementary to the single-stranded DNA over a large enough region to allow the two strands

to associate such that cleavage may occur at the chosen temperature and solely at the desired location.

Typically, the oligonucleotides of this preferred method of the invention are about 17 to about 5 30 nucleotides in length. Below about 17 bases, annealing is too weak and above 30 bases there can be a loss of specificity. A preferred length is 18 to 24 bases.

Oligonucleotides of this length need not be 10 identical complements of the germline genes. Rather, a few mismatches taken may be tolerated. Preferably, however, no more than 1-3 mismatches are allowed. Such mismatches do not adversely affect annealing of the oligonucleotide to the single-stranded DNA. Hence, the 15 two DNAs are said to be functionally complementary.

The second method to manipulate the amplified single-stranded DNAs of this invention for cloning comprises the steps of:

(i) contacting the nucleic acid with a 20 partially double-stranded oligonucleotide, the single-stranded region of the oligonucleotide being functionally complementary to the nucleic acid in the region in which cleavage is desired, and the 25 double-stranded region of the oligonucleotide having a Type II-S restriction endonuclease recognition site, whose cleavage site is located at a known distance from the recognition site; and

30 (ii) cleaving the nucleic acid solely at the cleavage site formed by the complementation of the nucleic acid and the single-stranded region of the oligonucleotide;

the contacting and the cleaving steps being performed at a temperature sufficient to maintain the nucleic acid in substantially single-stranded form, the oligonucleotide being functionally complementary to the nucleic acid over a large enough region to allow the two strands to associate such that cleavage may occur at the chosen temperature and at the desired location, and the cleavage being carried out using a restriction endonuclease that is active at the chosen temperature.

10 This second method employs Universal Restriction Endonucleases ("URE"). UREs are partially double-stranded oligonucleotides. The single-stranded portion or overlap of the URE consists of a DNA adapter that is functionally complementary to the sequence to be cleaved in the single-stranded DNA. The double-stranded portion consists of a type II-S restriction endonuclease recognition site.

15 The URE method of this invention is specific and precise and can tolerate some (e.g., 1-3) mismatches in the complementary regions, i.e., it is functionally complementary to that region. Further, conditions under which the URE is used can be adjusted so that most of the genes that are amplified can be cut, reducing bias in the library produced from those genes.

20 The sequence of the single-stranded DNA adapter or overlap portion of the URE typically consists of about 14-22 bases. However, longer or shorter adapters may be used. The size depends on the ability of the adapter to associate with its functional complement in the single-stranded DNA and the temperature used for contacting the URE and the single-stranded DNA at the temperature used for cleaving the DNA with the type II-S enzyme. The adapter must be

functionally complementary to the single-stranded DNA over a large enough region to allow the two strands to associate such that the cleavage may occur at the chosen temperature and at the desired location. We
5 prefer single-stranded or overlap portions of 14-17 bases in length, and more preferably 18-20 bases in length.

The site chosen for cleavage using the URE is preferably one that is substantially conserved in the
10 family of amplified DNAs. As compared to the first cleavage method of this invention, these sites do not need to be endonuclease recognition sites. However, like the first method, the sites chosen can be synthetic rather than existing in the native DNA. Such
15 sites may be chosen by references to the sequences of known antibodies or other families of genes. For example, the sequences of many germline genes are reported at <http://www.mrc-cpe.cam.ac.uk/imt-doc/restricted/ok.html>. For example, one preferred
20 site occurs near the end of FR3 -- codon 89 through the second base of codon 93. CDR3 begins at codon 95.

The sequences of 79 human heavy-chain genes are also available at
<http://www.ncbi.nlm.nih.gov/entre2/nucleotide.html>.
25 This site can be used to identify appropriate sequences for URE cleavage according to the methods of this invention. See, e.g., Table 8B.

Most preferably, one or more sequences are identified using these sites or other available
30 sequence information. These sequences together are present in a substantial fraction of the amplified DNAs. For example, multiple sequences could be used to allow for known diversity in germline genes or for frequent somatic mutations. Synthetic degenerate

sequences could also be used. Preferably, a sequence(s) that occurs in at least 65% of genes examined with no more than 2-3 mismatches is chosen

URE single-stranded adapters or overlaps are then made to be complementary to the chosen regions. Conditions for using the UREs are determined empirically. These conditions should allow cleavage of DNA that contains the functionally complementary sequences with no more than 2 or 3 mismatches but that do not allow cleavage of DNA lacking such sequences.

As described above, the double-stranded portion of the URE includes a Type II-S endonuclease recognition site. Any Type II-S enzyme that is active at a temperature necessary to maintain the single-stranded DNA substantially in that form and to allow the single-stranded DNA adapter portion of the URE to anneal long enough to the single-stranded DNA to permit cleavage at the desired site may be used.

The preferred Type II-S enzymes for use in the URE methods of this invention provide asymmetrical cleavage of the single-stranded DNA. Among these are the enzymes listed in Table 800. The most preferred Type II-S enzyme is FokI.

When the preferred Fok I containing URE is used, several conditions are preferably used to effect cleavage:

- 1) Excess of the URE over target DNA should be present to activate the enzyme. URE present only in equimolar amounts to the target DNA would yield poor cleavage of ssDNA because the amount of active enzyme available would be limiting.
- 2) An activator may be used to activate part of the FokI enzyme to dimerize without causing

cleavage. Examples of appropriate activators are shown in Table 510.

- 3) The cleavage reaction is performed at a temperature between 45°-75°C, preferably above 50°C and most preferably above 55°C.

The UREs used in the prior art contained a 14-base single-stranded segment, a 10-base stem (containing a FokI site), followed by the palindrome of the 10-base stem. While such UREs may be used in the methods of this invention, the preferred UREs of this invention also include a segment of three to eight bases (a loop) between the FokI restriction endonuclease recognition site containing segments. In the preferred embodiment, the stem (containing the FokI site) and its palindrome are also longer than 10 bases. Preferably, they are 10-14 bases in length. Examples of these "lollipop" URE adapters are shown in Table 5.

One example of using a URE to cleave an single-stranded DNA involves the FR3 region of human heavy chain. Table 508 shows an analysis of 840 full-length mature human heavy chains with the URE recognition sequences shown. The vast majority (718/840=0.85) will be recognized with 2 or fewer mismatches using five UREs (VHS881-1.1, VHS881-1.2, VHS881-2.1, VHS881-4.1, and VHS881-9.1). Each has a 20-base adaptor sequence to complement the germline gene, a ten-base stem segment containing a *FokI* site, a five base loop, and the reverse complement of the first stem segment. Annealing those adapters, alone or in combination, to single-stranded antisense heavy chain DNA and treating with *FokI* in the presence of, e.g., the activator FOKIact, will lead to cleavage of the antisense strand at the position indicated.

Another example of using a URE(s) to cleave a single-stranded DNA involves the FR1 region of the human Kappa light chains. Table 512 shows an analysis of 182 full-length human kappa chains for matching by the four 19-base probe sequences shown. Ninety-six percent of the sequences match one of the probes with 2 or fewer mismatches. The URE adapters shown in Table 512 are for cleavage of the sense strand of kappa chains. Thus, the adaptor sequences are the reverse complement of the germline gene sequences. The URE consists of a ten-base stem, a five base loop, the reverse complement of the stem and the complementation sequence. The loop shown here is TTGTT, but other sequences could be used. Its function is to interrupt the palindrome of the stems so that formation of a lollypop monomer is favored over dimerization. Table 512 also shows where the sense strand is cleaved.

Another example of using a URE to cleave a single-stranded DNA involves the human lambda light chain. Table 515 shows analysis of 128 human lambda light chains for matching the four 19-base probes shown. With three or fewer mismatches, 88 of 128 (69%) of the chains match one of the probes. Table 515 also shows URE adapters corresponding to these probes. Annealing these adapters to upper-strand ssDNA of lambda chains and treatment with *FokI* in the presence of FOKIact at a temperature at or above 45°C will lead to specific and precise cleavage of the chains.

The conditions under which the short oligonucleotide sequences of the first method and the UREs of the second method are contacted with the single-stranded DNAs may be empirically determined. The conditions must be such that the single-stranded DNA remains in substantially single-stranded form.

More particularly, the conditions must be such that the single-stranded DNA does not form loops that may interfere with its association with the oligonucleotide sequence or the URE or that may themselves provide
5 sites for cleavage by the chosen restriction endonuclease.

The effectiveness and specificity of short oligonucleotides (first method) and UREs (second method) can be adjusted by controlling the
10 concentrations of the URE adapters/oligonucleotides and substrate DNA, the temperature, the pH, the concentration of metal ions, the ionic strength, the concentration of chaotropes (such as urea and formamide), the concentration of the restriction
15 endonuclease(e.g., *FokI*), and the time of the digestion. These conditions can be optimized with synthetic oligonucleotides having: 1) target germline gene sequences, 2) mutated target gene sequences, or 3) somewhat related non-target sequences. The goal is to
20 cleave most of the target sequences and minimal amounts of non-targets.

In the preferred embodiment of this invention, the single-stranded DNA is maintained in substantially that form using a temperature between
25 45°C to 75°C. More preferably, a temperature between 50°C and 60°C, most preferably between 55°C and 60°C, is used. These temperatures are employed both when contacting the DNA with the oligonucleotide or URE and when cleaving the DNA using the methods of this
30 invention.

The two cleavage methods of this invention have several advantages. The first method allows the individual members of the family of single-stranded DNAs to be cleaved solely at one substantially

conserved endonuclease recognition site. The method also does not require an endonuclease recognition site to be built in to the reverse transcription or amplification primers. Any native or synthetic site in
5 the family can be used.

The second method has both of these advantages. In addition, the URE method allows the single-stranded DNAs to be cleaved at positions where no endonuclease recognition site naturally occurs or
10 has been synthetically constructed.

Most importantly, both cleavage methods permit the use of 5' and 3' primers so as to maximize diversity and then cleavage to remove unwanted or deleterious sequences before cloning and display.

15 After cleavage of the amplified DNAs using one of the methods of this invention, the DNA is prepared for cloning. This is done by using a partially duplexed synthetic DNA adapter, whose terminal sequence is based on the specific cleavage
20 site at which the amplified DNA has been cleaved.

The synthetic DNA is designed such that when it is ligated to the cleaved single-stranded DNA, it allows that DNA to be expressed in the correct reading frame so as to display the desired peptide, polypeptide
25 or protein on the surface of the genetic package. Preferably, the double-stranded portion of the adapter comprises the sequence of several codons that encode the amino acid sequence characteristic of the family of peptides, polypeptides or proteins up to the cleavage
30 site. For human heavy chains, the amino acids of the 3-23 framework are preferably used to provide the sequences required for expression of the cleaved DNA.

Preferably, the double-stranded portion of the adapter is about 12 to 100 bases in length. More

preferably, about 20 to 100 bases are used. The double-standard region of the adapter also preferably contains at least one endonuclease recognition site useful for cloning the DNA into a suitable display
5 vector (or a recipient vector used to archive the diversity). This endonuclease restriction site may be native to the germline gene sequences used to extend the DNA sequence. It may be also constructed using degenerate sequences to the native germline gene
10 sequences. Or, it may be wholly synthetic.

The single-stranded portion of the adapter is complementary to the region of the cleavage in the single-stranded DNA. The overlap can be from about 2 bases up to about 15 bases. The longer the overlap,
15 the more efficient the ligation is likely to be. A preferred length for the overlap is 7 to 10. This allows some mismatches in the region so that diversity in this region may be captured.

The single-stranded region or overlap of the
20 partially duplexed adapter is advantageous because it allows DNA cleaved at the chosen site, but not other fragments to be captured. Such fragments would contaminate the library with genes encoding sequences that will not fold into proper antibodies and are
25 likely to be non-specifically sticky.

One illustration of the use of a partially duplexed adaptor in the methods of this invention involves ligating such adaptor to a human FR3 region that has been cleaved, as described above, at 5'-ACnGT-
30 3' using HpyCH4III, Bst4CI or TaaI.

Table 250 F.2 shows the bottom strand of the double-stranded portion of the adaptor for ligation to the cleaved bottom-strand DNA. Since the HpyCH4III-Site is so far to the right (as shown in Table 206), a

sequence that includes the *Afl*III-site as well as the
XbaI site can be added. This bottom strand portion of
the partially-duplexed adaptor, H43.XAExt,
incorporates both XbaI and *Afl*III-sites. The top strand
5 of the double-stranded portion of the adaptor has
neither site (due to planned mismatches in the segments
opposite the XbaI and *Afl*III-Sites of H43.XAExt), but
will anneal very tightly to H43.XAExt. H43AExt
contains only the *Afl*III-site and is to be used with the
10 top strands H43.ABr1 and H43.ABr2 (which have
intentional alterations to destroy the *Afl*III-site).

After ligation, the desired, captured DNA can
be PCR amplified again, if desired, using in the
preferred embodiment a primer to the downstream
15 constant region of the antibody gene and a primer to
part of the double-standard region of the adapter. The
primers may also carry restriction endonuclease sites
for use in cloning the amplified DNA.

After ligation, and perhaps amplification, of
20 the partially double-stranded adapter to the single-
stranded amplified DNA, the composite DNA is cleaved at
chosen 5' and 3' endonuclease recognition sites.

The cleavage sites useful for cloning depend
on the phage or phagemid into which the cassette will
25 be inserted and the available sites in the antibody
genes. Table 1 provides restriction endonuclease data
for 75 human light chains. Table 2 shows corresponding
data for 79 human heavy chains. In each Table, the
endonucleases are ordered by increasing frequency of
30 cutting. In these Tables, Nch is the number of chains
cut by the enzyme and Ns is the number of sites (some
chains have more than one site).

From this analysis, *Sfi*I, *Not*I, *Afl*III, *Apa*LI, and *Asc*I are very suitable. *Sfi*I and *Not*I are preferably used in pCES1 to insert the heavy-chain display segment. *Apa*LI and *Asc*I are preferably used in
5 pCES1 to insert the light-chain display segment.

*Bst*EII-sites occur in 97% of germ-line JH genes. In rearranged V genes, only 54/79 (68%) of heavy-chain genes contain a *Bst*EII-Site and 7/61 of these contain two sites. Thus, 47/79 (59%) contain a
10 single *Bst*EII-Site. An alternative to using *Bst*EII is to cleave via UREs at the end of JH and ligate to a synthetic oligonucleotide that encodes part of CH1.

One example of preparing a family of DNA sequences using the methods of this invention involves
15 capturing human CDR 3 diversity. As described above, mRNAs from various autoimmune patients is reverse transcribed into lower strand cDNA. After the top strand RNA is degraded, the lower strand is immobilized and a short oligonucleotide used to cleave the cDNA
20 upstream of CDR3. A partially duplexed synthetic DNA adapter is then annealed to the DNA and the DNA is amplified using a primer to the adapter and a primer to the constant region (after FR4). The DNA is then
25 cleaved using *Bst*EII (in FR4) and a restriction endonuclease appropriate to the partially double-stranded adapter (e.g., *Xba* I and *Afl*III (in FR3)). The DNA is then ligated into a synthetic VH skeleton such as 3-23.

One example of preparing a single-stranded
30 DNA that was cleaved using the URE method involves the human Kappa chain. The cleavage site in the sense strand of this chain is depicted in Table 512. The

oligonucleotide kapextURE is annealed to the
oligonucleotides (kaBR01UR, kaBR02UR, kaBR03UR, and
kaBR04UR) to form a partially duplex DNA. This DNA is
then ligated to the cleaved soluble kappa chains. The
ligation product is then amplified using primers
kapextUREPCR and CKForeAsc (which inserts a AscI site
after the end of C kappa). This product is then
cleaved with ApaLI and AscI and ligated to similarly
cut recipient vector.

Another example involves the cleavage
illustrated in Table 515. After cleavage, an extender
(ON_LamEx133) and four bridge oligonucleotides (ON_LamB1-
133, ON_LamB2-133, ON_LamB3-133, and ON_LamB4-133) are
annealed to form a partially duplex DNA. That DNA is
ligated to the cleaved lambda-chain sense strands.
After ligation, the DNA is amplified with ON_Lam133PCR
and a forward primer specific to the lambda constant
domain, such as CL2ForeAsc or CL7ForeAsc (Table 130).

In human heavy chains, one can cleave almost
all genes in FR4 (downstream, i.e. toward the 3' end of
the sense strand, of CDR3) at a BstEII-Site that occurs
at a constant position in a very large fraction of
human heavy-chain V genes. One then needs a site in
FR3, if only CDR3 diversity is to be captured, in FR2,
if CDR2 and CDR3 diversity is wanted, or in FR1, if all
the CDR diversity is wanted. These sites are
preferably inserted as part of the partially double-
stranded adaptor.

The preferred process of this invention is to
provide recipient vectors having sites that allow
cloning of either light or heavy chains. Such vectors
are well known and widely used in the art. A preferred
phage display vector in accordance with this invention

is phage MALIA3. This displays in gene III. The sequence of the phage MALIA3 is shown in Table 120A (annotated) and Table 120B (condensed).

The DNA encoding the selected regions of the light or heavy chains can be transferred to the vectors using endonucleases that cut either light or heavy chains only very rarely. For example, light chains may be captured with *Apa*LI and *Asc*I. Heavy-chain genes are preferably cloned into a recipient vector having *Sfi*I, *Nco*I, *Xba*I, *Afl*III, *Bst*EII, *Apa*I, and *Not*I sites. The light chains are preferably moved into the library as *Apa*LI-*Asc*I fragments. The heavy chains are preferably moved into the library as *Sfi*I-*Not*I fragments.

Most preferably, the display is had on the surface of a derivative of M13 phage. The most preferred vector contains all the genes of M13, an antibiotic resistance gene, and the display cassette. The preferred vector is provided with restriction sites that allow introduction and excision of members of the diverse family of genes, as cassettes. The preferred vector is stable against rearrangement under the growth conditions used to amplify phage.

In another embodiment of this invention, the diversity captured by the methods of the present invention may be displayed in a phagemid vector (e.g., pCES1) that displays the peptide, polypeptide or protein on the III protein. Such vectors may also be used to store the diversity for subsequent display using other vectors or phage.

In another embodiment, the mode of display may be through a short linker to three possible anchor domains. One anchor domain being the final portion of M13 III ("IIIstump"), a second anchor being the full

length III mature protein, and the third being the M13 VIII mature protein.

The IIIstump fragment contains enough of M13 III to assemble into phage but not the domains involved in mediating infectivity. Because the w.t. III and VIII proteins are present, the phage is unlikely to delete the antibody genes and phage that do delete these segments receive only a very small growth advantage. For each of the anchor domains, the DNA encodes the w.t. AA sequence, but differs from the w.t. DNA sequence to a very high extent. This will greatly reduce the potential for homologous recombination between the display anchor and the w.t. gene that is also present.

Most preferably, the present invention uses a complete phage carrying an antibiotic-resistance gene (such as an ampicillin-resistance gene) and the display cassette. Because the w.t. *iii* and *viii* genes are present, the w.t. proteins are also present. The display cassette is transcribed from a regulatable promoter (e.g., P_{LacZ}). Use of a regulatable promoter allows control of the ratio of the fusion display gene to the corresponding w.t. coat protein. This ratio determines the average number of copies of the display fusion per phage (or phagemid) particle.

Another aspect of the invention is a method of displaying peptides, polypeptides or proteins (and particularly Fabs) on filamentous phage. In the most preferred embodiment this method displays FABS and comprises:

- a) obtaining a cassette capturing a diversity of segments of DNA encoding the elements:

$P_{reg}::RBS1::SS1::VL::CL::stop::RBS2::SS2::VH::CH1::$

linker::anchor::stop::,

where P_{reg} is a regulatable promoter, RBS1 is a first
ribosome binding site, SS1 is a signal sequence
5 operable in the host strain, VL is a member of a
diverse set of light-chain variable regions, CL is a
light-chain constant region, stop is one or more stop
codons, RBS2 is a second ribosome binding site, SS2 is
a second signal sequence operable in the host strain,
10 VH is a member of a diverse set of heavy-chain variable
regions, CH1 is an antibody heavy-chain first constant
domain, linker is a sequence of amino acids of one to
about 50 residues, anchor is a protein that will
assemble into the filamentous phage particle and stop
15 is a second example of one or more stop codons; and
b) positioning that cassette within the phage
genome to maximize the viability of the phage
and to minimize the potential for deletion of
the cassette or parts thereof.

20

The DNA encoding the anchor protein in the
above preferred cassette should be designed to encode
the same (or a closely related) amino acid sequence as
is found in one of the coat proteins of the phage, but
25 with a distinct DNA sequence. This is to prevent
unwanted homologous recombination with the w.t. gene.
In addition, the cassette should be placed in the
intergenic region. The positioning and orientation of
the display cassette can influence the behavior of the
30 phage.

In one embodiment of the invention, a
transcription terminator may be placed after the second
stop of the display cassette above (e.g., Trp). This
will reduce interaction between the display cassette

and other genes in the phage antibody display vector (PADV).

In another embodiment of the methods of this invention, the phage or phagemid can display proteins
5 other than Fab, by replacing the Fab portions indicated above, with other protein genes.

Various hosts can be used for growth of the display phage or phagemids of this invention. Such hosts are well known in the art. In the preferred
10 embodiment, where Fabs are being displayed, the preferred host should grow at 30°C and be RecA⁻ (to reduce unwanted genetic recombination) and EndA⁻ (to make recovery of RF DNA easier). It is also preferred that the host strain be easily transformed by
15 electroporation.

XL1-Blue MRF' satisfies most of these preferences, but does not grow well at 30°C. XL1-Blue MRF' does grow slowly at 38°C and thus is an acceptable host. TG-1 is also an acceptable host although it is
20 RecA⁺ and EndA⁺. XL1-Blue MRF' is more preferred for the intermediate host used to accumulate diversity prior to final construction of the library.

After display, the libraries of this invention may be screened using well known and
25 conventionally used techniques. The selected peptides, polypeptides or proteins may then be used to treat disease. Generally, the peptides, polypeptides or proteins for use in therapy or in pharmaceutical compositions are produced by isolating the DNA encoding
30 the desired peptide, polypeptide or protein from the member of the library selected. That DNA is then used in conventional methods to produce the peptide, polypeptides or protein it encodes in appropriate host cells, preferably mammalian host cells, e.g., CHO

cells. After isolation, the peptide, polypeptide or protein is used alone or with pharmaceutically acceptable compositions in therapy to treat disease.

EXAMPLES

5 **Example 1: Capturing kappa chains with BsmAI:**

A repertoire of human-kappa chain mRNAs was prepared by treating total or poly(A+) RNA isolated from a collection of patients having various autoimmune diseases with calf intestinal phosphatase to remove the
10 5'-phosphate from all molecules that have them, such as ribosomal RNA, fragmented mRNA, tRNA and genomic DNA. Full length mRNA (containing a protective 7-methyl cap structure) is unaffected. The RNA is then treated with tobacco acid pyrophosphatase to remove the cap
15 structure from full length mRNAs leaving a 5'-monophosphate group.

Full length mRNA's were modified with an adaptor at the 5' end and then reversed transcribed and amplified using the GeneRACE™ method and kit
20 (Invitrogen). A 5' biotinylated primer complementary to the adaptor and a 3' primer complementary to a portion of the construct region were used.

Approximately 2 micrograms (ug) of human kappa-chain (Igkappa) gene RACE material with biotin
25 attached to 5'-end of upper strand was immobilized on 200 microliters (μL) of Seradyn magnetic beads. The lower strand was removed by washing the DNA with 2 aliquots 200 μL of 0.1 M NaOH (pH 13) for 3 minutes for the first aliquot followed by 30 seconds for the second
30 aliquot. The beads were neutralized with 200 μL of 10 mM Tris (pH 7.5) 100 mM NaCl. The short oligonucleotides shown in Table 525 were added in 40

fold molar excess in 100 μ L of NEB buffer 2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol pH 7.9) to the dry beads. The mixture was incubated at 95°C for 5 minutes then cooled down to 55°C over 30
5 minutes. Excess oligonucleotide was washed away with 2 washes of NEB buffer 3 (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol pH 7.9). Ten units of BsmAI (NEB) were added in NEB buffer 3 and incubated for 1 h at 55°C. The cleaved downstream DNA was
10 collected and purified over a Qiagen PCR purification column (FIGs. 3 and 4).

A partially double-stranded adaptor was prepared using the oligonucleotide shown in Table 525. The adaptor was added to the single-stranded DNA in 100
15 fold molar excess along with 1000 units of T4 DNA ligase (NEB) and incubated overnight at 16°C. The excess oligonucleotide was removed with a Qiagen PCR purification column. The ligated material was amplified by PCR using the primers kapPCRT1 and kapfor
20 shown in Table 525 for 10 cycles with the program shown in Table 530.

The soluble PCR product was run on a gel and showed a band of approximately 700 n, as expected (FIGs. 5 and 6). The DNA was cleaved with enzymes
25 ApaLI and AscI, gel purified, and ligated to similarly cleaved vector pCES1. The presence of the correct size insert was checked by PCR in several clones as shown in FIG. 15.

Table 500 shows the DNA sequence of a kappa
30 light chain captured by this procedure. Table 501 shows a second sequence captured by this procedure. The closest bridge sequence was complementary to the sequence 5'-agccacc-3', but the sequence captured reads

5'-Tgccacc-3', showing that some mismatch in the overlapped region is tolerated.

Example 2: Construction of Synthetic CDR1 and CDR2 Diversity in V-3-23 VH Framework

5 A synthetic Complementary Determinant Region (CDR) 1 and 2 diversity was constructed in the 3-23 VH framework in a two step process: first, a vector containing the 3-23 VH framework was constructed, and then, a synthetic CDR 1 and 2 was assembled and cloned
10 into this vector.

 For construction of the V3-23 framework, 8 oligos and two PCR primers (long oligonucleotides: TOPFR1A, BOTFR1B, BOTFR2, BOTFR3, F06, BOTFR4, ON-vgC1, and ON-vgC2 and primers: SFPRMET and BOTPCRPRIM, shown in
15 Table 600) that overlap were designed based on the Genebank sequence of V323 VH. The design incorporated at least one useful restriction site in each framework region, as shown in Table 600. In Table 600, the segments that were synthesized are shown as bold, the
20 overlapping regions are underscored, and the PCR priming regions at each end are underscored. A mixture of these 8 oligos was combined at a final concentration of 2.5uM in a 20ul Polymerase Chain Reaction (PCR) reaction. The PCR mixture contained 200uM dNTPs, 2.5mM
25 MgCl₂, 0.02U *Pfu Turbo*TM DNA Polymerase, 1U Qiagen HotStart Taq DNA Polymerase, and 1X Qiagen PCR buffer. The PCR program consisted of 10 cycles of 94°C for 30s, 55°C for 30s, and 72°C for 30s. The assembled V3-23 DNA sequence was then amplified, using 2.5ul of a 10-
30 fold dilution from the initial PCR in 100ul PCR reaction. The PCR reaction contained 200uM dNTPs, 2.5mM MgCl₂, 0.02U *Pfu Turbo*TM DNA Polymerase, 1U Qiagen

HotStart Taq DNA Polymerase, 1X Qiagen PCR Buffer and 2 outside primers (SFPRMET and BOTPCRPRIM) at a concentration of 1uM. The PCR program consisted of 23 cycles at 94°C for 30s, 55°C for 30s, and 72°C for 60s.

5 The V3-23 VH DNA sequence was digested and cloned into pCES1 (phagemid vector) using the *SfiI* and *BstEII* restriction endonuclease sites (All restriction enzymes mentioned herein were supplied by New England BioLabs, Beverly, MA and used as per manufacturer's
10 instructions).

Stuffer sequences (shown in Table 610 and Table 620) were introduced into pCES1 to replace CDR1/CDR2 sequences (900 bases between *BspEI* and *XbaI* RE sites) and CDR3 sequences (358 bases between *AflII*
15 and *BstEII*), prior to cloning the CDR1/CDR2 diversity. The new vector is pCES5 and its sequence is given in Table 620. Having stuffers in place of the CDRs avoids the risk that a parental sequence would be over-represented in the library. The CDR1-2 stuffer
20 contains restriction sites for *BglII*, *Bsu36I*, *BclI*, *XcmI*, *MluI*, *PvuII*, *HpaI*, and *HincII*, the underscored sites being unique within the vector pCES5. The stuffer that replaces CDR3 contains the unique restriction endonuclease site *RsrII*. The stuffer
25 sequences are fragments from the penicillase gene of *E. coli*.

For the construction of the CDR1 and CDR2 diversity, 4 overlapping oligonucleotides (ON-vgC1, ON_Br12, ON_CD2Xba, and ON-vgC2, shown in Table 600
30 and Table 630) encoding CDR1/2, plus flanking regions, were designed. A mix of these 4 oligos was combined at a final concentration of 2.5uM in a 40ul PCR reaction. Two of the 4 oligos contained variegated sequences

positioned at the CDR1 and the CDR2. The PCR mixture contained 200uM dNTPs, 2.5U Pwo DNA Polymerase (Roche), and 1X Pwo PCR buffer with 2mM MgSO₄. The PCR program consisted of 10 cycles at 94°C for 30s, 60°C for 30s, and 72°C for 60s. This assembled CDR1/2 DNA sequence was amplified, using 2.5ul of the mixture in 100ul PCR reaction. The PCR reaction contained 200uM dNTPs, 2.5U Pwo DNA Polymerase, 1X Pwo PCR Buffer with 2mM MgSO₄ and 2 outside primers at a concentration of 1uM. The PCR program consisted of 10 cycles at 94°C for 30s, 60°C for 30s, and 72°C for 60s. These variegated sequences were digested and cloned into the V3-23 framework in place of the CDR1/2 stuffer.

We obtained approximately 7×10^7 independent transformants. Into this diversity, we can clone CDR3 diversity either from donor populations or from synthetic DNA.

It will be understood that the foregoing is only illustrative of the principles of this invention and that various modifications can be made by those skilled in the art without departing from the scope of and spirit of the invention.